Twitter 上のメッセージによる国のイメージ測定：内容分析とテキストマイニングによる分析

著者	石井 健一
発行年	2012-06
その他のタイトル	The Measurement of Nation Images Based on Content Analysis and Text-Mining of Twitter Messages
シリーズ	Department of Social Systems and Management Discussion Paper Series;no.1294
URL	http://hdl.handle.net/2241/117462

Department of Social Systems and Management

Discussion Paper Series

No.1294

Twitter 上のメッセージによる国のイメージ測定

―内容分析とテキストマイニングによる分析

(The Measurement of Nation Images Based on Content Analysis and Text-Mining of Twitter Messages)

by

石井 健一
(Kenichi ISHII)

June 2012

UNIVERSITY OF TSUKUBA
Tsukuba, Ibaraki 305-8573
JAPAN

Twitter 上のメッセージによる国のイメージ測定

―内容分析とテキストマイニングによる分析[i]

石井健一

要約　本報告は、国のイメージの測定に向けて行なったいくつかの試行的な測定結果を報告するものである。API を通じて収集した Twitter のメッセージを内容分析とテキストマイニングで分析し、さらにブログ検索のヒット件数で国と文化キーワードとの距離を計算した。いくつかの結果は、中国のイメージがきわめて悪いことを示していた。ただし、インターネット上のテキストを用いて国のイメージを測定することについて、方法論的な問題がなお残されていることも確認された。

研究の背景と目的

最近 SNS の利用率が急速に増加している。SNS で最も利用度が高いといわれる Twitter は API を公開しており、メッセージをプログラムから得ることが可能である[ii]。Twitter は日本人利用者の場合、身近な友だちとのものが比較的多いとみられ、日本人の本音の会話に近いものが多く含まれていると予想される(石井 2011a)。本研究は、Twitter のメッセージを利用し、日本人の対外イメージを分析することを目的とする。

本報告書は、科学研究費補助金基盤（B)『日中の相互国家イメージと「国家ブランディング」の可能性―中国と日本での実証研究』(代表石井健一)において実施したいくつかの分析結果を報告するものである。本研究では、Twitter から得られたメッセージに対して二種類の分析を行なう。一つは評定者が内容を読んでコーディングする内容分析である。もう一つは、テキストマイニングによる分析である。また、これらとは別にブログの検索数を用いて国のイメージを測定した結果も報告する。

方法

(1)Twitter メッセージの内容分析

API のプログラムをつくり 2011 年 9 月 14 日から 10 月 27 日にかけて Twitter で「中国」「台湾」「香港」「韓国」「シンガポール」「マレーシア」「インド」「アメリカ」「イタリア」「ドイツ」「フランス」「ロシア」を含むメッセージを収集した。

ただし、「中国」は日本の「中国地方」の意味でも使われているので、中国関係のメッセージは一名のコーダーがすべてのメッセージをチェックして、国名としての意味以外で使われている場合はデータから削除した。その結果、分析に用いたメッセージの総件数は、

44,024 件である（ただし、一つのメッセージが複数の国のデータに入っている重複も少数だがある)。

そのうち、3,057 件のメッセージを無作為に抽出し、二人のコーダーで分担して、どういう側面のメッセージか(経済・技術等、政治・軍事、国民性、その国の道徳、対日感情)をコーディングしたうえで、その国の日本に対する総合的な評価を(1)肯定的、(2)否定的、(3)どちらでもないの三段階で評価してもらった(この場合の「肯定的」とは、日本や日本人にとって肯定的という意味である)。ただし、メッセージを読んでその国と全く関係のない話題の場合はデータから除外したところ(たとえば「フランスパン」についてのメッセージなど)、国のメッセージとして評価対象となったのは 2,560 件であった。

なお、3,057 件のうち、1,686 件については二名の評定者が独立に評定し、1371 件については、一名の評定者が単独で評価した。二人の評定者の一致度をスピアマンの順位相関係数で測定したところ r=0.328 (p<.001)であった。二人の評定が一致しない場合は、どちらかの評定をランダムに選んだ。

(2) Twitter メッセージのテキストマイニング

上記の 44,024 件のメッセージを対象として R という統計ソフト上で MeCab というプログラムを用いて形態素解析を行い分析した(石田 2008)。

表 1　ブログ検索にもちいたキーワード

	日本語	英語		日本語	英語
1	日本	Japan	13	ブラジル	Brazil
2	アメリカ	US	14	オーストラリア	Australia
3	中国	China	15	音楽	Music
4	韓国	South Korea	16	芸術	Art
5	台湾	Taiwan	17	ファッション	Fashion
6	イタリア	Italy	18	観光	Tour
7	フランス	France	19	グルメ	gourmet
8	ドイツ	Germany	20	スポーツ	Sports
9	イギリス	UK	21	歌手	Singer
10	インド	India	22	アニメ	Animation
11	ロシア	Russia	23	映画	Movie
12	インドネシア	Indonesia	24	文学	Literature

(3)　ブログ検索を用いた国のイメージ測定

グーグルには「ブログ検索」というサービスがあり、ブログの中から特定のキーワードをもつブログ数を表示することができる[iii]。表 1 のキーワードを用いてそれらを二つずつ組み合わせた場合のヒット件数を測定した。測定は、日本語と英語について別々に行なった。

このヒット件数に基づいて、キーワードiとキーワードjの距離行列Dijを以下のように定義した(Fiはキーワードjの単独でのヒット件数であり、Fijはキーワードiとjを組み合わせた時のヒット件数である)。

$$Dij = Fij/(Fi \times Fj)$$

なお、表1のキーワードは、1-14が主要国の国名、15-24が文化関係の単語となっている。

結果

(1)内容分析による分析結果

評定者によるメッセージの評価の結果は、表1のようになった。ここでは、ポジティブ・ネガティブのいずれにしろ評価のあるメッセージの比率を「関心度」、評価のあるメッセージのうちポジティブなメッセージの比率を「好意度」として各国の値を計算した。

その結果、好意度が最も低いのが中国であり、次いでアメリカ、ロシア、韓国といった国が低かった。中国については、各種の世論調査の結果にそう結果であるが、アメリカは一般には最も日本人が高い好意度をもっているとされている国の一つであり、意外な結果といえる。韓国はネガティブなメッセージが二番目に多いが、ポジティブなメッセージも多い点が注目される。中国は好意度は最も低いが、関心度は平均くらいであまり高いとは言えない。

上で定義した「好意度」と内閣府が毎年調査している「外交に関する調査」の「親しみを感じる」%の値とのスピアマンの順位相関係数を求めたところ、値は0.04ときわめて低かった。この原因としてアメリカに対する評価が「外交に関する調査」ではきわめてよいのに対して、Twitter上のメッセージでは好意度が中国に次いで低かったことある。

メッセージ内容を(1)経済・技術、(2)政治・軍事、(3)国民性やその国の道徳性に関するものの三つに分類した結果が表2である(これ以外のメッセージは「それ以外」とむなる)。ここで注目されるのは、韓国に関しては国民性や道徳に関するメッセージがきわめて多いということである。

表２　各国のメッセージの内容分析の結果

	総件数（T）	ポジティブなメッセージの件数（A）	ネガティブなメッセージの件数（B）	ポジティブなメッセージの比率（好意度）A/(A+B)	評価のあるメッセージの比率（関心度）(A+B)/T	外交に関する世論調査「親しみを感じる」%[iv]
中国	229	1	27	3.6	12.2	26
台湾	215	51	1	98.1	24.2	
香港	231	24	5	82.8	12.6	
韓国	204	27	42	39.1	33.8	62
シンガポール	222	17	5	77.3	9.9	51
マレーシア	227	13	3	81.3	7.0	51
インド	191	11	12	47.8	12.0	41
アメリカ	186	14	50	21.9	34.4	82
イタリア	208	20	7	74.1	13.0	64
ドイツ	258	6	2	75.0	3.1	64
フランス	206	17	8	68.0	12.1	64
ロシア	183	5	14	26.3	10.4	13

表３　各国に関するメッセージの内容ジャンル(件数)

	(1)経済・技術等の側面	(2)政治・軍事の側面	(3)国民性、その国の道徳、対日感情の側面	(3)の%
中国	18	28	12	21
台湾	2	52	16	23
香港	3	29	29	48
韓国	6	69	148	66
シンガポール	4	22	7	21
マレーシア	3	16	11	37
インド	10	23	5	13
アメリカ	10	64	22	23
イタリア	17	27	8	15
ドイツ	9	8	5	23
フランス	12	25	6	14
ロシア	4	19	27	54
計	98	382	296	38

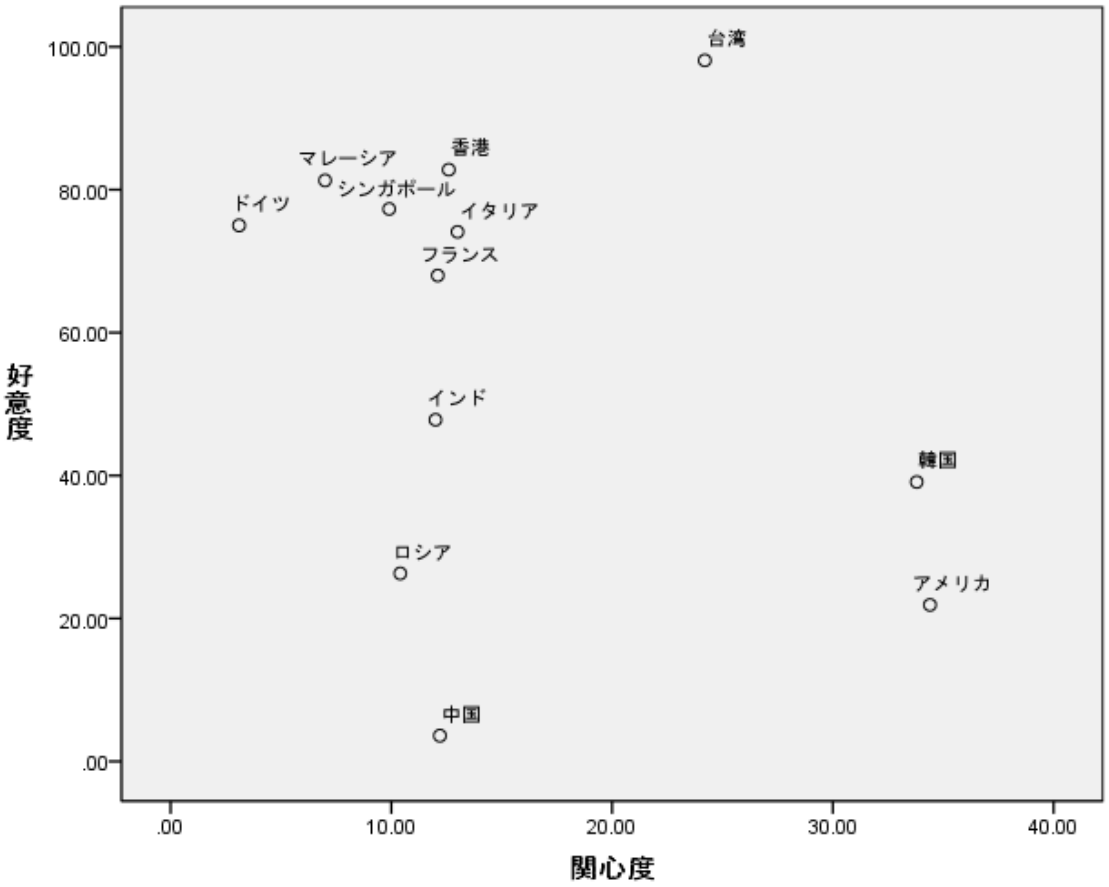

図１　メッセージにおける各国の好意度と関心度

(2) テキストマイニングによる分析結果

次に Twitter で抽出したメッセージをテキストマイニングで分析した結果を報告する。本分析では、12 カ国の国名で検索した結果得られた総計 44024 件のメッセージを分析対象とする。

表 4　テキストマイニングで対象としたメッセージの件数

国	件数	国	件数
中国	3567	インド	3254
台湾	3867	アメリカ	3854
香港	3882	イタリア	3619
韓国	4022	ドイツ	3771
シンガポール	3649	フランス	3718
マレーシア	3599	ロシア	3222

次にこれらの国名がどのような名詞と形容詞と結びつきやすいのかを $\chi 2$ 値を用いて調べてみた。なお、まず出現頻度が合計で 10 回以上という条件で少数回しか登場しない単語は省いた。まず χ^2 の大きな単語をとりだし、各単語について最も頻度が多いものをその国の「特徴語」とした。

名詞

表 5 には、χ^2 値と頻度が 10 以上という基準にあう各国のメッセージの特徴語の上位をリストにしたものである。上位に入っているリストをみると、台湾で「義捐金」、韓国で「民主党」「慰安婦」「竹島」、ドイツで「原発」「福島」と言った時事ニュースに関連する単語がいくつか入っていることがわかる。また、中国で「烏龍茶」、韓国で「キムチ」、インドで「カレー」、イタリアで「パスタ」、ドイツで「ビール」、フランスで「パン」など有名な食品名は上位に登場している。

形容詞

表 6 には $\chi 2$ が 3 以上のものを掲載したが名詞に比べて $\chi 2$ の値は大きくない。概して特徴的な形容詞はあまりないと言えそうである。中国の「青い」は、これを含むメッセージが何回かリツイートされていたためである。また、これらは関係するメッセージの中に出現したというだけであり、必ずしも当該国を修飾する用語として使われていたわけではないことにも注意が必要である。

表５　各国の特徴語（名詞のみ。数値はχ^2値）

中国		台湾		香港		韓国		シンガポール		マレーシア	
米	5639	義援金	9391	発	7567	@	110766	アジア	6685	ニュース	16376
国	3304	日本人	8161	海外	5951	日本	85082	調達	1387	KL	13813
漢	2067	ー	7034	円	5259	人	27865	日経	1128	QUOTE	13499
政府	2054	億	6456	旅行	3426	韓	7285	移転	749	位	10857
経済	1502	拡散	5718	中	2899	ー	3619	航空	724	タイ	6842
米国	1363	希望	5463	万	2763	者	2985	到着	701	ベトナム	3137
中国人	1081	報道	4894	欠航	2401	民主党	2400	ホテル	683	企業	2280
署名	1049	動画	4774	予約	2326	笑	1683	市場	483	参加	1782
茶	997	親日	4523	ジョブズ	2151	♪	1495	取引	466	インドネシ	1740
北京	815	地震	4087	今日	2088	女性	1449	先物	455	戦	961
産経	651	さ	3981	月	2040	問題	1309	物流	418	輸出	956
式	651	メディア	3908	SIM	2038	流	1182	パナソニ	406	クアラルン	944
アリ	585	たち	3674	ツアー	1866	在日	1121	本部	406	カ	742
産	578	九	3527	学生	1799	ドラマ	1056	所	394	チリ	709
会長	521	半年	3435	着	1596	兆	905	カジノ	321	選手	697
国家	515	ﾀﾞﾝﾄﾂ	3401	映画	1591	朝鮮	826	進出	267	等	672
MSN	488	思い	3401	空港	1576	gt	778	早め	261	r	671
軍	477	突破	3390	日間	1569	T	669	戦略	245	フィリピン	631
ババ	458	分の	3337	年	1532	慰安	649	開催	232	blog	593
烏龍茶	415	州	3333	上海	1476	婦	645	明日	221	ブルネイ	577
佛	391	比べ	3296	行	1462	竹島	638	平均	218	シモンチ	563
チベット	355	収入	3272	投資	1344	フジテレビ	597	先	205	工場	560
不動産	326	ネット	3160	iPhone	1312	好き	577	広島	188	雇用	547
技術	316	人達	2927	格安	1295	アイドル	527	kotarotar	180	加盟	490
省	310	これ	2750	版	1203	野田	475	便	162	MotoGP	464
春	310	私	1869	追悼	1173	神	408	あと	154	GP	424
買収	300	Po	1673	氏	1074	ファン	392	新卒	147	生産	405
お願い	294	Pon	1671	ロゴ	984	勉強	381	来年	143	現地	403
関心	291	marumar	1476	フリー	964	禁止	379	Record	128	仕事	391
開発	285	観光	1439	デザイン	934	気	372	機能	126	事業	362
可能	283	公式	1411	国際	856	お祝い	367	バンコク	123	トレイン	351
上	246	ちゃん	1240	現在	818	抗議	346	枚	120	ブルー	345
事件	230	麻生	1083	東京	744	キムチ	345	セミナー	119	オーストラ	339
SMAP	227	素敵	1035	評判	698	通貨	334	政策	117	ペルー	332
ロイター	225	Taiwan	950	マカオ	685	スワップ	330	studySing	115	ウォール	332
員	213	杯	901	出発	647	ソウル	328	就職	112	関税	296
パンダ	212	大学生	798	行き	635	北	322	友達	111	お供	282
非常	207	支援	798	w	619	反日	318	SGX	109	開始	272
		制作	764	権	535	発売	313	人材	108	調査	270
		Beautiful	763			ウォン	310	前日	105	崩壊	266
		女子	753			外国	297	削減	103	続き	258
								移住	100		
								チャンギ	98		

表 5 の続き

インド		アメリカ		イタリア		ドイツ		フランス		ロシア	
カレー	11106	RT	86801	日	10874	語	64726	パン	10136	金髪	2367
世界	3798	の	26967	スペイン	6262	原発	12050	さん	5112	東欧	2164
F	3403	こと	15686	教育	4540	衛星	4300	事故	3106	大統領	2159
GP	2475	ん	13178	料理	3951	福島	4160	仏	1785	秘密	2055
タブレット	2313	TPP	11296	格下げ	3749	サッカー	3107	放射線	1522	PC	2049
店	973	デモ	2989	パスタ	3664	落下	2357	パリ	1490	本気	2001
何	972	的	2851	電車	1641	ビール	2098	国民	1266	サービス	1886
台	807	よう	2394	ユーロ	1624	人工	1564	IRSN	1252	娘	1844
安	594	ため	2311	ギリシャ	1413	ZDF	923	性	1221	分	1571
率	542	金	2079	ヨーロッパ	1228	労働	876	原子力	1194	プーチン	1213
ドル	541	手	1735	国債	1074	テレビ	851	欧州	1175	北方領土	1139
成長	484	今	1487	銀行	1067	実態	813	イギリス	1055	via	839
sasakitos	445	それ	1405	ローマ	969	日本語	673	東電	1022	tubemani	637
RNM	415	事	1267	段階	930	リーグ	653	ベクレル	956	ｽﾃｷ	625
屋	363	基準	1198	危機	866	番組	616	フランス語	872	TNp	565
農民	344	そう	1153	債務	760	aya	575	発表	751	Wjl	565
自由	340	もの	1099	格付け	709	製	546	水	688	nhk	522
遺伝子	322	Bq	1069	遺産	680	横浜	466	安全	669	首相	515
モンサント	311	TPP	1066	機	617	NHK	453	セシウム	636	news	499
自殺	282	時	957	懸念	614	独	450	食品	630	健康	492
数	266	家	909	展	561	出場	445	放射	621	領	466
建築	264	無料	828	意味	560	会社	432	販売	602	mig	465
千	253	保険	676	Wikipedia	481	センター	406	派	592	ヨ	425
すべて	239	購入	634	ピザ	437	ビジネス	343	倍	587	政権	403
News	239	英語	623	ワイン	431	市議	323	研究所	562	後	400
社	231	自分	579	ムーディー	414	人々	300	君	530	汚染	394
系	210	政治	537	雨	395	学校	298	防護	488	北朝鮮	365
普及	206	マン	532	定刻	373	量	297	過去	469	ソ連	332
生徒	192	みたい	487	南	355	輸入	293	いち	469	モスクワ	305
価格	190	男性	481	代表	277	午前	292	最大	460	雪男	290
必要	185	歳	461	終了	276	市民	285	京	452	宇宙	269
ワタ	181	社会	458	波及	266	放射能	282	核	448	新潟	248
音楽	172	キャプテン	457	送料	251	字幕	270	S	447	笑顔	235
種	163	化	444	空気	236	香川	259	monjukur	441	廃墟	220
携帯	152	医療	431	ミラノ	217	監督	252	おもらし	441	除	219
ビザ	152	交渉	400	速報	190	子供	247	割	438	カラパイア	216
端末	150	手当	392	見通し	189	チェルノブ	244	革命	428	美女	205
配付	148	話	362	伊	178	値	238	冷却	417		
ガンジー	147	カダフィ	358	不安	177	犯罪	237	海	391		
精神	145	代	344	気持ち	174	市長	234	施設	350		
製造	143	層	332	悪化	167	地球	231	ラ	338		
mp	131	歴史	326	物	166	今度	221	朝	337		
公開	129	反対	315	身	163	確率	218	俺	327		
ヶ月	128	戦争	262	起源	159	放送	213				
		昨日	257	名前	157	エネルギ	211				
				名	151	作業	207				
						城	200				

表 6　各国の形容詞の特徴語（χ^2>3）

中国		台湾		香港		韓国		シンガポール		マレーシア	
青い	148.0	かわいい	190.8	甘い	23.4	楽しい	193.1	低い	153.0	美味しい	266.3
珍しい	73.8	っぽい	134.1	かっこいい	13.7	早い	190.4	凄い	130.9	蒸し暑い	70.0
面白い	64.9	おいしい	108.1	黒い	13.2	暑い	72.7	明るい	49.4	美しい	41.4
やすい	58.9	おかしい	103.2	よろしい	3.6	新しい	39.1	貧しい	44.4	優しい	20.3
永い	24.6	ひどい	65.6	かっこよい	3.4	遅い	28.6	望ましい	27.4	重い	18.1
危ない	15.4	うまい	47.2	ふさわしい	3.2	深い	17.8	美味い	9.0	うれしい	7.0
上手い	14.6	ものすごい	21.9			懐かしい	12.5	濃い	8.2	難い	3.6
弱い	14.1	もの凄い	21.7			涼しい	7.9	温かい	7.6		
小さい	10.6	宜しい	15.3			熱い	6.1	興味深い	7.1		
こわい	5.1	速い	10.6			軽い	4.3	すい	5.7		
激しい	4.7	白い	8.4			有り難い	4.0	ええ	3.2		
怪しい	3.2	ずるい	7.7					めんどくさ	3.1		
		生々しい	6.1								
		くい	5.5								
		情けない	5.2								
		にくい	4.9								
		でかい	3.2								

インド		アメリカ		イタリア		ドイツ		フランス		ロシア	
良い	327.9	安い	541.5	おそい	1113.7	厳しい	67.8	ない	8856.9	いい	2938.1
無い	164.3	高い	379.5	怖い	168.3	偉い	48.7	悲しい	3192.3	多い	456.8
悪い	142.3	恐ろしい	20.5	若い	85.2	眠い	37.2	大きい	3177.3	よい	283.7
欲しい	94.2	忙しい	18.6	親しい	24.2	おもしろい	26.1	ひもじい	405.4	広い	171.1
ほしい	89.5	悔しい	4.4	暗い	22.5	羨ましい	23.6	可愛い	356.0	すごい	133.2
嬉しい	54.3			近い	21.6	難しい	23.1	寒い	160.3	素晴らしい	127.6
酷い	53.7					赤い	11.3	恥ずかしい	70.1	少ない	85.8
辛い	32.1					めざましい	10.9	暖かい	4.7	強い	72.6
遠い	16.9					イイ	7.4	古い	3.9	長い	38.2
ぽい	16.7					短い	4.7	うい	3.4	正しい	24.7
痛い	15.5									カッコイイ	9.3
やばい	13.9									厚い	3.7
詳しい	10.4										
うらやまし	6.2										
極まりない	5.1										
さむい	4.2										
さみしい	4.1										
寂しい	3.5										
仲良い	3.4										

(3) ブログ検索による分析結果

ブログ検索で得られた距離行列を用いて R の計量多次元尺度分析を行なった。2 次元までの解を求めた結果が表 7 である。

日本語と英語ともに第一軸で「中国」の値が突出して高くなっている。一方、負の値には日本語・英語ともに文化関係の単語が集まっている(音楽、映画など)。これは、中国を語るブログがこうした文化関係のことを語らないことを示すものといえる。

第二軸は、日本語では「日本」の値が最も低く、「ファッション」の値が最も高くなって

いる。英語では、「イタリア」「インドネシア」の値が低く、「イギリス」や「音楽」の値が高くなっている。第二軸の解釈は困難である。

表 7　ブログ検索から得られた結果の多次元尺度分析の結果

	日本語		英語	
	第 1 軸	第 2 軸	第 1 軸	第 2 軸
日本	0.0	−2.2	0.8	−0.1
アメリカ	0.4	−0.2	0.0	−1.1
中国	2.7	−0.1	7.6	0.0
韓国	0.5	−0.2	−0.3	−0.3
台湾	0.1	0.8	−0.4	0.2
イタリア	0.1	0.0	1.4	−2.6
フランス	0.1	0.0	0.7	−0.2
ドイツ	0.1	0.0	0.9	−1.2
イギリス	0.1	−0.1	3.9	3.7
インド	0.1	−0.1	2.6	−0.6
ロシア	0.0	−0.1	−0.4	−0.8
インドネシア	0.1	−0.1	1.6	−2.6
ブラジル	0.1	−0.2	−0.1	−0.8
オーストラリア	0.1	−0.1	1.3	−0.2
音楽	−1.7	0.0	−0.9	2.0
芸術	0.0	0.1	−0.9	1.7
ファッション	0.1	1.8	−1.3	0.5
観光	0.2	0.2	−2.3	0.4
グルメ	0.2	0.5	−1.3	1.0
スポーツ	0.2	−0.2	−0.6	1.2
歌手	−0.5	0.5	−0.9	−1.0
アニメ	−1.1	0.0	−0.9	1.0
映画	−1.9	−0.1	−9.9	−0.3
文学	0.0	−0.1	−0.6	0.1

次にブログ検索で得られた距離行列を用いて、各国と文化に関する単語との距離の和を計算してみた。その結果が表 8 である。日本語ブログの場合は、ロシア、イタリア、インドネシアといった国との距離が短くなっていて、これらの国について Twitter で語る場合は文化に関する話題が多いことを示唆している。英語ブログの場合は、台湾、韓国、ロシア、

ブラジルという順に距離が短くなっている。また、日本語・英語いずれの場合も、中国との距離が最も遠くなっていて、中国がブログで語られる時は、こうした文化に関するキーワードが使われない傾向があることを示している。

表 8　文化キーワードとの距離

	日本語ブログ	英語ブログ
日本	17.6	27.8
アメリカ	12.3	39.3
中国	22.4	50.4
韓国	10.9	20.7
台湾	6.3	18.1
イタリア	5.1	36.7
フランス	6.3	30.2
ドイツ	6.0	37.2
イギリス	5.1	58.6
インド	5.4	41.5
ロシア	4.4	23.9
インドネシア	5.6	48.7
ブラジル	6.2	23.9
オーストラリア	6.0	35.3

結論

本分析は方法論を含めて試行的なものである。Twitter 上のメッセージを分析した結果は、以下のような問題点が明らかになった。第一は、すべてのメッセージが当該国に対してポジティブまたはネガティブな態度を明確にしているわけではないことである。Twitter のメッセージの多くは個人的なやりとりであり、国について論じる内容は必ずしも頻度としては多くはない。したがって、内容分析にせよ、テキストマイニングにせよ、この規模でのメッセージの分析から明確に態度の方向性を見出すことは難しいようである。

また、いくつかの Twitter メッセージはリツイートされて複数回データの中に出現していた。リツイートには短いコメントなどが書き込まれることもあり、リツイートが同一のメッセージとは言えないが、このように繰り返して引用されるメッセージの比重が高くなってしまうことは否定できない。

ブログ検索の方法はさらに試行的なものである。しかし、ブログという大きな単位でキーワードが共起することが二つのキーワードの類似性を必ずしも意味しないかもしれない。

今回、分析で得られた結果の多くも解釈が困難であった。また、Google が表示するヒット件数についてもどのようなアルゴリズムによるものなのかが分からないという問題もある。ブログ検索による測定は、簡易で便利な方法ではあるが、方法論的な問題が残されているといえよう。

参考文献

石田基広 (2008) R によるテキストマイニング入門、森北出版

石井健一 (2011a) マイクロブログ Twitter における日本人利用者の特徴 Department of Social Systems and Management Discussion Paper Series, no.1277

石井健一 (2011b) Facebook と Twitter の発言における特徴語の比較 Department of Social Systems and Management Discussion Paper Series, no.1279

---

[i] 本調査は平成 23 年度科学研究費補助金基盤（B)『日中の相互国家イメージと「国家ブランディング」の可能性―中国と日本での実証研究』（代表石井健一）による。

[ii] ただし、全てのメッセージのデータが得られるわけではない。

[iii] http://www.google.co.jp/blogsearch?hl=ja (日本語)
http://www.google.com/blogsearch?hl=en (英語)である。

[iv] 平成 23 年に実施された内閣府の「外交に関する調査」の「親近感を感じる」(全体)の%。ただし、イタリア、フランス、ドイツは「ヨーロッパ諸国」、シンガポールとマレーシアは「東南アジア諸国」に対する親近感の値を用いている。
http://www8.cao.go.jp/survey/h23/h23-gaiko/index.html